POLICY RESEARCH WORKING PAPER 1506

Restructuring Regulation of the Rail Industry for the Public Interest

Ioannis N. Kessides
Robert D. Willig

To a greater extent than in the past, market forces should shape the prices and logistics of railroad services. Noncommercial services, where needed, should be based on an explicit agreement between the railway and the government, one that views public service obligations as a business relationship between a customer (government) and the supplier (railway).

The World Bank
Private Sector Development Department
September 1995

POLICY RESEARCH WORKING PAPER 1506

Summary findings

Throughout the world, the rail industry historically has been one of the most extensively regulated of all sectors. Price, entry, exit, financial structure, accounting methods, vertical relations, and operating rules have all been subject to some form of government control. The public utility paradigm of government regulation has been applied on the assumption that the economic characteristics of the rail industry preclude competitive organization or the need for market responsiveness.

In the past three decades, however, policymakers and economists have become increasingly critical of traditional regulation of the rail industry. It is generally accepted that in markets where rail carriers seek to meet demand, there is often effective competition, and that government restrictions on the structure and conduct of firms in this industry impose considerable costs on society.

Misguided regulatory policies have been blamed for the misallocation of freight traffic among competing modes of transport, excess capacity, excessive operating costs, and poor investment decisions. Regulatory controls have also shouldered much of the blame for the poor financial condition of railroads, the deterioration of rail plant, the suppression and delay of cost-reducing innovations, and the mediocre quality of rail service.

Kessides and Willig suggest principles for restructuring railroad regulation — indeed, for restructuring the orientation of railroad entities — for the sake of the public interest.

Much can be learned, they contend, from applying the principles of industrial organization to analysis of the rail industry. To assess the implications of policies aimed at rate regulation or infrastructure investment, it is essential to understand the nature of technology, costs, and demand in the rail industry. Government's role in relation to market behavior should be based explicitly on the economic and technological realities of the railroad marketplace.

Kessides and Willig say that restructuring along the lines they suggest — putting more emphasis on marketing effectiveness — will result in a more profitable railway with a better chance of covering its costs for commercial services. Changing the basis for noncommercial services as they suggest will make those services more effective at fulfilling public policy objectives, will eliminate an insuperable drain on revenues that condemns rails to inadequate investment, and will eliminate cross-subsidies that make it difficult for rails to compete against other modes of transport.

This paper — a product of the Private Sector Development Department — is part of a larger effort in the department to analyze the regulatory issues in network utilities. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Jocelyn Dytang, room G4-027, telephone 202-473-7161, fax 202-522-3181, Internet address jdytang@worldbank.org. September 1995. (45 pages)

RESTRUCTURING REGULATION OF THE RAIL INDUSTRY FOR THE PUBLIC INTEREST

Ioannis N. Kessides, The World Bank

and

Robert D. Willig, Princeton University

TABLE OF CONTENTS

1. INTRODUCTION

Throughout the world, the rail industry historically has been one of the most extensively regulated sectors in the economy.[1] Price, entry, exit, financial structure, accounting methods, vertical relations and operating rules have all been subject to some form of governmental control. The application of the public utility paradigm of governmental regulation has been expressly premised on the assumption that the economic characteristics of the rail industry preclude competitive organization and any need for market responsiveness.

Over the past three decades, however, economists and policy makers have become increasingly critical of the traditional public utility regulation of the rail industry.[2] It is generally understood that in the relevant economic markets in which rail carriers seek to meet demand, there is often (but surely not always) effective competition. It is generally agreed that governmental restrictions upon the structure and conduct of firms in this industry impose considerable costs upon society. Misallocation of freight traffic among competing transport modes, excess capacity, excessive operating costs, and poor investment decisions are often the result of misguided regulatory policies. Regulatory controls have, therefore, been held responsible in large part for the poor financial condition of the railroads, for the deterioration of the rail plant, for the suppression and delay of cost-reducing innovations, and for the mediocre quality of rail service.

The purpose of this paper is to suggest a set of principles for restructuring railroad regulation, and indeed for restructuring the orientation of railroad entities, for the sake of the public interest. Our methodology is to first focus on the economic characteristics of the rail industry and their implications for the design of efficient regulatory policy. Then, we apply powerful sets of analytic tools, in the context of the industry characteristics, to clarify the relevant principles for reform. One of the main points that emerges from the application of industrial organization analysis is that much can be learned about actual and potential industry structure and performance, and about policies designed to promote the public interest, from pertinent understanding of market demands for the industry's products and the nature of the productive techniques available to the industry's firms. Indeed, before we can fully assess the implications of policies aimed at rate regulation or infrastructure investments, it is essential to have a full understanding of the nature of technology, costs, and demand facing the rail industry. The role of the government in relation to market behavior should, therefore, be explicitly based on the underlying economic characteristics of the industry and the technological conditions of its production.

---

[1] See Friedlaender (1969) and Keeler (1983).

[2] See among others: Friedlaender (1971); Levin (1978, 1981); and Beyer (1987).

We hope to impart the following message in the course of this chapter: recent developments in industrial organization analysis as well as in regulatory practice call for a major reorientation of public policy towards railroads, one that follows a set of principles we shall articulate.

2. CURRENT ISSUES OF PUBLIC POLICY IN THE RAIL INDUSTRY[3]

The economic characteristics of the rail industry make it a natural target for government intervention, yet also render it particularly difficult to regulate in the public interest. The old regulatory systems failed to handle the central regulatory problem arising in railroads and certain other major industries (e.g., telecommunications, electric power, and postal services): the mixture of competition and monopoly elements in supply. Indeed, in these industries, just as in the railroad industry over the years, regulation has stifled competition in the provision of services, restricted the benefits of economies of scope, retarded innovation, fostered inefficient service, and thereby harmed the public interest, while at the same time protecting it from the exploitation of monopoly power. The first-best lesson of the perfect competition model, calling for prices to be set equal to marginal costs, has no doubt contributed to the common regulatory ethos which *seeks to equate* price to *some* measure of cost. This doctrine has been used frequently where it is completely inappropriate and without logical foundation, that is, in cases where prices should be based on demand as well as cost considerations.

This section focuses on the central pricing issues involved in partial deregulation of railroad rates. It articulates principles to guide regulatory oversight of the rate setting of unsubsidized railroads--principles that are consistent with economic analysis and that are essential for protection of the public interest. Public interest regulatory oversight of railroad pricing involves two basic issues. The first of these is the adequacy of revenues, the determination of the criteria by which this can be judged, and the means by which it can be achieved. The second issue is the choice of rates that are consistent with adequate revenues and that are best for the public interest.

In a regime of deregulation, one of the key elements in protecting the public interest is the avoidance of any residual regulation which effectively prevents the achievement of financial viability by the rail network. The public will hardly be well served by a set of regulatory rules which condemn the railroads to an inability to compete in the financial marketplace and which, consequently, will result in a rail network which is increasingly obsolete, is characterized by deterioration, and in which cumulative abandonment of service becomes the guiding principle.

---

[3] It is a pleasure and an honor for the authors to acknowledge their debt to the thinking and written testimony of William J. Baumol on many of the subjects covered in this portion of the paper. A partial summary of this material can be found in Willig and Baumol (1987).

In determining prices for the outputs of multiproduct railroad firms, regulators have long faced a number of difficult issues that flow inexorably from the basic economic characteristics of the industry that we discussed above. The endemic economies of scale and scope imply that straightforward measures of costs cannot be used to dictate pricing. Economies of scale imply that marginal cost pricing, absent subsidy to the firm or multipart tariffs, will not allow the firm to break even. Further, the shared costs that are a concomitant of economies of scope cannot be unambiguously identified with individual products, so that any rule selected to associate shared costs with individual services will be arbitrary. Such arbitrary measures as fully distributed (or "fully allocated") costs, therefore, cannot substitute for marginal cost measures as decision rules for proper pricing, and the search for any purely cost-based substitute rule is a remnant of inappropriate reliance on the model of perfect competition for guidance on regulation.

A system of rate regulation based upon fully distributed costs, where costs are apportioned on any basis other than demand, is inappropriate because prices set by that method are highly unlikely to permit railroads to achieve an adequate rate of return. Moreover, such a method leads to serious inefficiency by discouraging innovation and by generating prices that are too high to attract competitive traffic, which severely restrict the amount of services delivered by railroads, and which thus produce still higher rates for the remaining traffic.[4]

By contrast, there are sound pricing principles which promote economic efficiency while simultaneously removing impediments to adequate returns for carriers. These principles can be applied in practically useful fashions to assess the reasonableness of those rates which are judged to require continued regulatory oversight.[5] The principles lead to demand differentiated prices, sometimes referred to as Ramsey prices, which apportion all unattributable fixed and common costs of the railroad among its services on the basis of the values of those services to consumers, mathematically expressed as their elasticities of demand. By providing that each service is priced at a markup over marginal costs which is inversely related to the elasticity of demand for that service, economically efficient differential pricing combines cost and demand factors in an optimal manner. These principles result in lower prices for shippers generally by establishing a set of rates which encourages the purchase of more rail transportation services by more shippers than artificial fully distributed cost based pricing, thereby creating a larger traffic base over which unattributable costs can be apportioned. Ramsey pricing maximizes the opportunity for rail carriers to earn an adequate rate of return on capital, and they foster innovation and efficiency in the provision of rail transportation services by rewarding carriers who achieve cost reductions.

---

[4] See Braeutigam (1977), Kahn (1988).

[5] See Braeutigam (1979, 1984).

Economically efficient differential pricing is entirely consistent with the hallmark of deregulation: that market forces, rather than regulation, should control rates for transportation services. Thus, when a particular type of traffic is subject to competition, direct or indirect, regulatory intervention is unjustified because that competition will produce efficient prices without regulatory guidance. Furthermore, so long as a railroad's earnings fall short of its cost of capital, the need for regulatory constraints upon any of that carrier's rates is minimal and, to the extent such a constraint prevents the carrier from earning an adequate return in the future, it is contrary to the public interest. By definition, there is no danger that such a carrier is receiving excessive overall profits derived from market power or any other cause. In addition, if the rate for any service supplied by a railroad not yet earning adequate revenues overall is held down by regulation below that level which consumers of that service are prepared to pay rather than do without the service, then, in the long run, even those consumers will be harmed--the carrier will find it unprofitable to invest the necessary replacement and maintenance capital, causing a deterioration in, and ultimate withdrawal of, the service.

2.1 The Proper Criterion for Adequacy of Revenues

Since avoiding impairment of financial viability plays so crucial a role in any rational program of rate regulation, it is important to describe the criterion by which financial viability can be judged. Just what information is required to determine when a firm's revenues are adequate to cover its pertinent costs? While the answer would appear to be obvious, the past history of regulation demonstrates rather forcefully that it is in fact widely misunderstood. The basic issue is that among the costs which must always be included in these calculations is the cost of the firm's capital, including any capital it has generated internally.

The logic of this criterion is straightforward. Revenues are defined to be adequate when they are just sufficient to enable the firm to attract the capital needed for maintenance, replacement, modernization and whatever expansion demand conditions justify. If revenues are lower than this, the deterioration and eventual disappearance of the service in question are a foregone conclusion.

Adequate revenues are those which provide a rate of return on net investment equal to the current cost of capital (i.e., the level of return available on alternative investments). This is the revenue level necessary for a railroad to compete equally with other firms for available financing in order to maintain, replace, modernize, and, where appropriate, expand its facilities and services. If railroads cannot earn the fair market rate of return, their ability both to retain existing investments and obtain new capital will be impaired, because both the existing and prospective funds could be invested elsewhere at a more attractive rate of return. Indeed, the market for funds is one of the most competitive in the economy. It simply

offers no room to those who cannot meet the competition for funds by others who come there to seek capital. Thus, there is no escaping the following principles that determine the adequacy of revenues:

(a) The firm's overall rate of return must be equal to the returns currently earned by the typical firm with similar risks elsewhere in the economy. Otherwise the required funds will be denied to it.

(b) This means that adequacy of revenues can only be judged by comparison with earnings outside the regulated industry, not by comparing the regulated industry's earnings with the market value of its equity. For the market prices of those securities will automatically adjust themselves downward to match any act by the regulator which restricts the earnings of a firm below a compensatory rate of return, and so such a comparison will appear to justify any earnings restrictions, no matter how inappropriate.

(c) In determining the revenue requirements for financial viability, the rate of return obtained by comparison with other industries must be applied to a rate base which covers the economic replacement cost (under regulation) of all facilities. (Suitably updated historic costs may be utilized instead of replacement costs if the allowed rate is expressed in nominal terms).

(d) With the rate base determined in this way and the rate of return on that rate base equal to the cost of capital, as given by earnings prevailing elsewhere in the economy, one will have determined the figure for total net earnings by the railroad that can appropriately be considered to be adequate for it to compete successfully in the capital market.

(e) This earning figure must not be applied as a rigid ceiling. Otherwise railroads will not have the ability to earn this figure over the long run, since they will be precluded from making up for the revenue shortfalls which may occur as the result of temporary downward fluctuations in demand for their services.

For prices to make sense economically they must be never be incompatible with this earnings level. Of course, no prices can guarantee that a railroad will earn adequate returns overall. For if demands for its services are insufficient or the railroads' operations are conducted wastefully or its services are poor, even appropriate prices cannot be expected to lead to profitable operation. But once the railroads are permitted to charge appropriate prices in a competitive environment, the regulatory impediments to financial viability will have been cleared away. It is then up to the railroads to take advantage of the opportunity by means of economy of operation, quality of service and effective marketing effort.

2.2 The Regulatory Problem

Indivisibilities, pervasive economies of scale and scope, high costs of entry, and small-numbers competition in the railroad industry are all consistent with the likely persistence of prices in excess of marginal cost. However, while scale economies go hand-in-hand with natural monopoly, a railroad may or may not have the price-setting discretion that characterizes the textbook monopolist. It all depends on whether the activities characterized by economies of scale and scope are shielded from other sources of competition in the relevant market, and whether there are protective barriers to entry.

In the railroad industry, extensive capital sums must be sunk in way and structures and in a variety of ancillary facilities in order to create new rail lines. The sunk cost and longevity of railroad capital may suggest that the railroad industry is one in which contestability analysis cannot conceivably apply--these sunk costs generally suffice to deter entry by new rail lines. However, railroad services are far more contestable than these impediments to rail entry would suggest, because there are often strong competitive pressures from other modes of transportation--such as trucking and water carriage--on the rates charged for shipment of a wide variety of commodities.[6]

The basic patterns of railroad regulation, established many decades ago in wholly different market conditions, are simply obsolete. Their premise was that railroads had a collective monopoly, or near-monopoly, in land transport. This condition has disappeared long ago, if indeed it ever existed. Nearly every sphere of rail freight service now faces intense competition. Rival products and rival sources of supply (including trucks, barges, and alternative rail routes) are likely to impose effective competitive constraints upon many, if not most, rail activities. In those activities where there is no evidence that it holds a position of market dominance the railroad should be offered freedom in pricing. Still, there remain instances in which the competitive checks of intramodal, intermodal, geographic, and product competition are weak or nonexistent. There is an understandable apprehension that in such cases market forces may not be relied upon to prevent excessive pricing. The resulting monopoly power is the basic justification for regulation of rail rates and earnings, and defines the basic task with which regulation must grapple.

However, it must be emphasized before discussing the appropriate means to deal with this issue that, in practice, effective competition can assume a variety of subtle forms. Therefore, one must never proceed in haste to undermine the workings of the market through special intervention. Railroads do not

---

[6] It should be noted that in many instances the relevant competition is not just on the route involved in the rail movement, but also on alternative routes that offer economic substitute services for the shipper. For example, a manufacturer may find it equally desirable to ship output to two very different places for the purposes of sale, and will choose the option with the least expensive transportation.

face only the competition of trucks and barges. For example, oil and natural gas shipped by pipeline competes with coal shipped by rail and since coal shipment is profitable to the railroads, the competition of petroleum products limit the price they can charge for carrying coal. Also, the market served by one railroad may compete for the coal with a market served by another and this too can keep rates in line.

2.3 The Cost Allocation Problem

The presence of substantial economies of scale and scope in the railroad industry creates a number of problems for government regulation. Perhaps the most troubling is the fact that it is impossible to allocate, in any nonarbitrary way, a share of fixed and common costs to any one of a railroad's many activities. There is simply no way to subdivide those costs in a mechanical fashion that is unique and has any foundation in economic logic.

In practice, regulatory authorities historically have determined tariffs based on so-called fully distributed (or allocated) costs. Under this method regulators do (somehow) allocate shared production costs to individual services. Each service is then required to generate revenues which will cover all the costs associated with that service. Although it is often argued that there is no sound economic rationale for fully distributed cost (FDC) pricing, this practice obviously does have economic consequences.

Traditionally, regulatory proceedings have focused on three types of FDC rules. The first of these is the distribution of shared costs on the basis of a common measure of utilization, such as gross ton-miles. Under this FDC approach, which is termed the *relative output* method, shared costs are allocated in proportion to the number of units of output of each service. A second approach sometimes used is the allocation of shared costs in proportion to the costs that can be directly attributed to the various services. This *attributable cost* method has also been traditionally used by many unregulated firms in their allocation of overhead costs. A third scheme requires allocation of shared costs in proportion to the gross revenues generated by each service. This *gross revenue* approach, has been frequently used to allocate overhead costs between freight and passenger services.

In addition to costs that are directly attributable, a service may also be assigned a portion of those costs which cannot be clearly associated with any one service. Some administrative costs are shared by several services. Railroad track is used in the transport of many kinds of freight. Shared costs may, therefore, comprise a large portion of total costs. Thus, the method of allocating shared costs may significantly influence the rate which may be required for any particular service.

2.4 No System of Fully Allocated Costs Can Yield Economically Efficient Prices

FDC pricing rules suffer from several disabilities: (i) since fully distributed costs bear no direct relationship to marginal costs, there is no basis in economic efficiency for FDC pricing; (ii) on grounds of economic efficiency, it may sometimes be desirable to set a price for some service so that the revenues it generates do not cover its fully distributed costs; (iii) because the determination of fully distributed costs is arbitrary, there is no economic basis for concluding that a service is being subsidized by other services if its revenues are less than its fully distributed costs; and (iv) FDC pricing is anticompetitive since it prevents a supplier from offering a service at a proposed tariff less than an FDC price, particularly if the proposed tariff exceeds the marginal cost of providing the service. In addition, there is circular reasoning behind the FDC practice. Tariffs which are determined to be "appropriate" at a given time may depend on the existing levels of output or revenues, and these, in turn, depend on previous tariffs. Thus, fully distributed costs may depend on the acceptance of a prior tariff structure.

The most serious defect of fully distributed costs as a basis for rate determination is that they do not necessarily measure marginal cost responsibility in a causal sense. They do not measure by what amount costs would be increased if additional quantities of any particular service were taken, or by what amount costs would be reduced if the service were correspondingly curtailed--they are costs that are averaged by an arbitrary method. Also, being apportionments of historical costs, even when they do accurately reflect historical responsibility for the incurrence of these costs among the respective users, they do not provide a reliable measure of what will happen to costs in the future if particular portions of the business are expanded or dropped.

Another defect of the fully allocated cost criterion is its complete neglect of any demand data. Even if it is based on "relative use" as measured in tons or ton-miles, it cannot capture the role of demand, which economic analysis has shown to be vital in the choice of optimal prices. Even the best intentioned of fully allocated cost standards must employ some rigid criterion to allocate the portion of a railroad's total costs which are not directly attributable to any one of its services in particular. But no such fixed allocation criterion can possibly reflect the subtleties, fine structure and changes in patterns of demand for the railroads' services that are induced by external developments and clearly call for adjustments in its prices. This, of course, is true not only of a standard fully allocated cost approach, but of any rigid formula which bases future prices on cost data of the past, because it too can not take account of changes in demand.

It may seem paradoxical that fully allocated cost criteria, that are apparently designed to assure that all costs are covered by revenues, can in fact preclude rail carriers from achieving financial viability. The reason is that ceilings based on fully allocated costs are set so that unattributable costs are divided

in an arbitrary manner among all types of traffic. Then, for the these costs to be recovered, all types of traffic must actually move at the rates that include the arbitrary cost allocations. But traffic with transport value that is below average for its tons, ton-miles, or other allocator will not move by rail at those rates. That is, any service whose demand is insufficient to cover its allocated share of total cost at the fully allocated cost determined price will have a revenue shortfall which fully allocated cost ceilings will prevent other services from making up. Consequently, if the unattributable costs are substantial, and if the values of rail services vary substantially, then fully allocated cost rate ceilings will preclude attainment of adequate revenues.

The effects of fully allocated cost pricing on the efficiency of the utilization of transport resources are equally pernicious. In doing their best to earn adequate revenues despite the handicap imposed by fully allocated cost rate ceilings, rail carriers will be unable to preserve traffic whose value to the shipper exceeds its attributable cost, but which falls sufficiently far below fully allocated cost. True, in the absence of fully allocated cost regulation, any such traffic could contribute revenues that exceed the costs that it causes, and would provide social benefits greater than social costs. But with fully allocated cost rate ceilings, this traffic will reduce the net revenues of the rail carrier and will thus not be compensatory. The reason is that this traffic will be assigned its portion of unattributable costs on the basis of its tons and ton-miles or some other arbitrary allocator, thereby reducing the share of those costs allocated to other traffic with higher value, and consequently reducing the ceiling and the rates on that traffic.

Fully allocated cost rate ceilings may also stifle the incentives of railroads to innovate and compete. A rail carrier cannot be expected to invest in new facilities, in research and development, and in marketing activities designed to elicit new traffic if the financial gains from the new traffic are counterbalanced by induced decreases in the ceilings on the rates charged to pre-existing traffic. Similarly, a rail carrier could not be expected to compete for freight by offering low rates if the necessary markups were much below the arbitrary allocations of unattributable costs; if it did so, it would never earn adequate revenues because its gain from the low-rated traffic would be outweighed by the induced decrease in the ceilings applied to more highly rated traffic.

2.5 Long-Run Marginal Cost and Pricing Efficiency

The indivisibilities, economies of joint production, and high fixed costs which make small numbers competition in the railroad industry an inevitable consequence also render the traditional measure of static deadweight loss incomplete as a welfare indicator. A regime of marginal cost pricing would eliminate the deadweight loss. But marginal cost pricing is a questionable regulatory objective, since the railroads would incur substantial losses. If the regulator attempts to force rates to equal marginal costs, overall revenues will fall short of overall costs. Without subsidy, reduction of the short-run welfare loss to zero

would cause the long-run deterioration of the industry's capital stock. For rail systems that are characterized by scale economies, rates must generally lie above the costs economically attributable to individual services if revenues are to cover total costs.

It should also be noted that the use of long-run marginal cost to measure pricing efficiency frequently leads to misguided rules which could force the railroad into a pattern of behavior that is in conflict with the dictates of the market. Indeed, the rigid requirement that each rate always cover the long-run marginal cost of service is tantamount to a prescription of pricing inefficiency for railroads. Moreover, such a misguided decision would be likely to impose a heavy penalty upon the public because it would sometimes deprive the public of a valuable service at a price it is willing to pay--a price which also best serves the interests of the company--namely a price that lies between long-run and short-run marginal cost.

The role of a cost floor as a measure of efficiency is to determine whether the railroad would be better off without the traffic in question. There are two basic reasons why it will often be appropriate for a price to lie below the corresponding long-run marginal cost. First, investment decisions which were entirely rational and appropriate at the time they were made will in many cases subsequently be affected by unexpected developments. Such eventualities may cast a shadow over the future of the service which utilizes the investment. A railroad is always better off carrying any and all traffic that can cover its short-run avoidable costs and make some contribution to its fixed and common costs--the supplier earns more by providing the service than by abandoning it. The test of efficient pricing above short-run avoidable costs is whether the railroad is pricing in accordance with market demand. So long as the revenue inadequate railroad is charging profit-maximizing rates, it is necessarily pricing efficiently; if the price maximizes the service's contribution to company profits, clearly no other price conceivably can bring that service closer to being compensatory in the long run.

The second reason why efficient prices will often fall short of long-run marginal cost affects even services whose financial viability is absolutely clear. Whether a railroad will be able in the long-run to earn revenues that are sufficient to cover the replacement cost of a particular service or a group of services depends on the level of demand over time. The rail industry is strongly affected by business fluctuations in the economy, and demand for individual rail services and groups of services can and does vary widely over time. Even services whose financial viability is absolutely clear, will certainly encounter years in which business is good and other years in which business conditions are poor. In the less prosperous years, the firm's earnings will often fall short of long-run marginal cost because market conditions permit no alternative. Of course, the shortfall will then be made up during the prosperous periods. In this manner then, the firm will in the long-run meet its revenue requirements. But to insist that prices always cover long-run marginal costs is effectively to undermine the market pricing process, and, very likely, even

the viability of the service--it would clearly distort the intertemporal pattern of usage of the service and so reduce economic efficiency. In addition, innovation and improvements in operating efficiency over time could potentially reduce costs and enhance contribution. A rule that assumed assets would not be replaced simply because current revenues from a particular service were depressed, would remove any incentive or ability to respond to upswings in demand or improvements in efficiency which would otherwise permit the service to continue.

The long-run marginal cost should never be used mechanically as a rigid minimum cost floor in the pricing of a railroad that is already extant. At the same time, it should be emphasized that the long-run marginal cost cannot serve legitimately to establish the level of efficient pricing above short-run costs at any point in time. Instead, efficient rates will always have to be consistent with demand. This is true whether or not a railroad has market dominance over a particular service and whether or not it has achieved adequacy of revenues. The demand for each service always helps to determine the contribution that service should make to the railroad's overall costs, and that it should make if its behavior is to comport with the requirements of economic efficiency.

2.6 Economically Efficient Pricing

If there were no need for enterprises to be financially self-supporting, an ideally efficient allocation of society's resources would be brought about if the price of each good or service were equal to its marginal cost. At such prices, consumers elect to purchase all units of goods and services that yield them benefits larger than the costs of providing them. And, in response to such prices, consumers avoid purchasing units that yield them benefits smaller than the costs of providing them. As a result, the economy misses no opportunity to allocate resources to uses where they yield benefits greater than costs, and no resources are allocated to uses with benefits lower than costs.

In industries without substantial fixed costs, competition tends to result in prices which approximate marginal or incremental costs. However, in the railroad industry, the prevalence of large fixed and common costs make it impossible for the supply of rail services to become financially self-supporting with marginal cost pricing. The financial infeasibility of marginal cost pricing rules out any sensible mechanical or formula-based procedure for regulatory determination of rates. In particular, compensatory rates cannot be determined by the regulator on the basis of cost data alone since the financial viability of any price depends also on the quantity of rail services customers are willing to buy at that price. This is true because there is no correlation between demand considerations and any cost accounting convention.

Allocation of fixed and common costs in accord with any non-demand based apportionment rule will almost invariably produce inconsistencies with the patterns of shipper demands. Some rates will be too low, and consequently the railroad will receive less than the optimal contribution from those services. Other rates will be too high, so that the railroad will either earn less than the optimal contribution or no contribution at all. In short, in a multiproduct industry with uncongested fixed and common costs, the pricing of individual services on the basis of any cost allocation is contrary to the interests of both the operating entities and the shipping public. Rational determination of prices must be based on both cost and demand conditions--demand considerations as well as cost data must enter into decision making, in order to permit adequacy of revenues and achieve efficiency.

2.7 Demand-Based Differential Pricing

Non-demand based cost apportionment methods do not necessarily reflect the railroad's ability (or inability) to impose the assigned allocations and cover its costs. Thus, they frequently "over-assign" or "under-assign" the carrier's unattributable costs to particular services. If a carrier sought to apply FDC pricing to all its traffic, it would lose that portion of the traffic for which demand could not support the price assigned. In that event, the remaining shippers would be saddled with a larger portion of the carrier's unattributable costs since they would no longer share those costs with the lost traffic.

Ramsey prices apportion all unattributable fixed and common costs of the railroad among its services on the basis of their demand characteristics. Each service is priced at a mark-up over marginal cost which is inversely related to the elasticity of demand for that service--services whose demands are highly elastic are assigned prices that are very close to their marginal costs, while services whose demands are very inelastic are priced well above those costs. The magnitude of these mark-ups among all services must be sufficiently high to earn net revenues that cover fixed and common costs and, hence, achieve revenue adequacy.

The logic of this inverse elasticity rule and its implied allocation of unattributable costs is quite simple. The elasticity of demand provides a quantitative interpretation of the traditional concept, *value of service*, which has played an important role in public utility pricing. Consumers who place relatively high value on a service will have demands for it that are relatively inelastic, and vice-versa. For if a rise in price would lead to no significant reduction in quantity demanded (that is, if demand is *inelastic*), then the service must be worth at least the higher price to its consumers, that is, the value of the service must be high. Conversely, if a rise in the price of a service would lead consumers to curtail their demand substantially (that is, if demand is quite elastic), then the service must be worth little or no more to its consumers than the original price, so that the value of the service must be low.

In view of this correspondence between value of service and demand elasticity, the inverse elasticity rule of Ramsey pricing can be restated in terms of a familiar and long-used principle in railroad pricing. Services with relatively high values to their consumers should contribute relatively large net revenues to the coverage of unattributable, fixed and common costs. Thus, the implicit allocation of unattributable costs should be based on value of service, rather than any *pro rata* sharing or other arbitrary method. All factors that influence a rail carrier's elasticities of demand are relevant for the carrier's Ramsey prices. These factors may include the value of the commodity shipped, intermodal competition, intramodal competition, interport competition, and the substitutability of other commodities for the one shipped at its destination. Value of service is therefore properly construed as a market concept--it refers to the value of the rail carrier's service with all demand factors considered, and generally cannot be evaluated by such measures as the ratio of a commodity's price to its weight alone.

2.8 Ramsey Pricing--Efficiency and Equity

Under Ramsey pricing, it is the "non-marginal" portion of total costs (i.e., the total cost less the marginal cost of each service multiplied by the quantity of the service provided) that is apportioned on the basis of demand. Equivalently, it is the shortfall between total costs and the revenues that would accrue from pricing each service at the level of its marginal cost. In the presence of economies of scale, this shortfall is positive. Ramsey prices, therefore, deviate from marginal costs only to the extent necessary to provide adequate revenues--they permit the railroad to achieve the goal of revenue adequacy with the least sacrifice of economic welfare compared with marginal cost pricing.

Increases above marginal cost in the price of an elastic service causes much traffic to be lost--traffic that would generate net benefits because it is valued above the cost it causes. However, less traffic is lost when the price of an inelastic service is raised, and the traffic that is curtailed is the least valued portion. Consequently, when prices must be elevated above marginal costs to cover unattributable costs, it is economically efficient to increase prices of inelastic services more than prices of elastic ones. Such Ramsey prices are, on average, the lowest consistent with financial viability. As long as the price charged to the price-elastic service exceeds its incremental cost, then the service is contributing to the carrier's overhead costs. Thus, Ramsey pricing principles benefit all shippers by establishing a set of rates which encourage the purchase of more transportation services by more shippers than artificial prices based on fully distributed cost. Therefore, by creating a larger traffic base over which unattributable costs can be apportioned, Ramsey pricing also benefits the so called captive shippers. The expansion of rail traffic represents an increase in the flow of commodities to their markets at lower transportation costs. As a result, social productivity is enhanced, and more consumers can obtain more of the goods they desire at lower costs of supply.

Since Ramsey prices are based on the relative values of the different services, they may seem to approximate the solution of the profit-maximizing monopolist, sometimes loosely described as charging what the market will bear. However, it is only the firm's necessary costs, including the cost of capital, that are covered by Ramsey prices. Monopoly prices, on the other hand, are controlled by no such constraint. Ramsey prices, therefore, are very different both qualitatively and quantitatively from monopoly prices.

It should also be emphasized that Ramsey prices are equitable. First, they are non-discriminatory in the sense that services with similar economic characteristics have similar prices, whatever the commodities shipped, whatever the route and whatever the identity of the shipper. That is, two different services with the same elasticities of demand will be priced at the same percentage mark-ups above marginal costs. And, two different services with the same marginal costs and demand elasticities will bear identical Ramsey prices. Second, while the Ramsey prices of different services are different proportions of the services' marginal costs, the burdens from these necessary mark-ups that are borne by the consumers have roughly the same proportion to their respective values of service.

2.9 The Stand-Alone Cost Constraint

Ramsey pricing requires that both the marginal cost and the elasticity of demand be quantified for every movement in the carrier's system. That is all but impossible to do with any degree of accuracy. Thus, while the Ramsey formula is useful as a theoretical guideline for rate determination, its application would be administratively difficult and burdensome--the amount of data and the analysis required are overwhelming. The Ramsey pricing rule has also been criticized because it does not constrain the railroad's pricing of traffic over which it possesses market dominance and its consequent failure to protect captive shippers. In addition, although Ramsey pricing minimizes the static welfare cost of the revenue adequacy constraint, still output levels are less than they would be if rates were set at marginal costs. This results in economic inefficiency because the value of the lost output to the shipper is greater than the value of the resources saved by reducing output. Under these conditions, it may be feasible for the parties to negotiate a contract with incentive clauses, volume-sensitive pricing, or two-part pricing that will leave both parties better off than at the flat Ramsey price, and consequently be yet more desirable for the public interest.

The critical issue from the standpoint of efficiency is the criterion used to set the ceiling on rates where there is market dominance. As we have noted above, rate ceilings derived from fully distributed costs are inimical to the public interest. Economically rational ceilings are obtainable from the *stand-alone cost* (SAC). The stand-alone cost to any captive shipper or group of shippers who benefit from sharing joint and common costs, is the cost of serving that shipper or group of shippers alone, as if the shipper

or its group were isolated from the railroads' other customers. A rate calculated by the SAC methodology represents the theoretically maximum rate that a railroad could levy on shippers without substantial diversion of traffic to a hypothetical competing service. Thus, the SAC criterion serves as a surrogate for competition and leads to a simulated competitive price. The competing service could be a shipper providing service for itself or a third party competing with the incumbent railroad for the traffic. In either case, the SAC represents the minimum cost of a possibly hypothetical alternative to the service provided by the incumbent railroad.

2.10 Stand-Alone Costs Provide Appropriate Protection Against Excessive Rates

The stand-alone cost test rules out the possibility of abuse of monopoly power by enforcing a competitive standard upon railroad rates. The hallmark of monopoly power is the elevation of the price of a service above the costs at which competitors could provide that service. The stand-alone cost test makes that impossible and imposes the same ceilings on rates for any traffic over which the railroad is dominant that the market would impose if it were subject to either active or potential competition. In the long run, in contestable markets, no group of shippers would agree to pay more to a carrier for their transportation services than it would cost them to produce these services for themselves, or than it would cost a competitor to supply it to them. In the short run, a rail carrier facing either active or potential effective competition could not obtain revenues from a group of shippers that exceeded their stand-alone costs, because those shippers could then be profitably served by a competitor charging lower rates. Thus, the stand-alone cost test affords shippers the same protection that effective competition would provide.

Clearly, the stand-alone cost is unnecessary and inappropriate where there is competition--the price set by competitors (reflecting current costs of service) will set a market ceiling. If only potential competition exists, the regulatory test is still unnecessary because if the rates charged by the existing carrier exceeded stand-alone costs, that fact would constitute an invitation to entry by the potential competitors. However, for any shippers who are truly captive, in that the rail carrier faces no effective direct, indirect or potential competition for their freight, the stand-alone cost does provide an economically rational ceiling.

No regulatory ceiling is needed to act as a surrogate for active or potential competition from a mode that can operate through the market. In the presence of such a competing mode market pressures will enforce the stand-alone cost ceiling--since no one will be able to sell at any higher price. There is yet another consideration which reduces further the likelihood that it will be necessary for regulators to intervene, except on the rarest occasions, in order to enforce observance of stand-alone cost ceilings upon rates. This consideration stems logically from the very concept of stand-alone cost. For if the rates for any service exceed those necessary to cover stand-alone cost, that fact by itself constitutes an invitation

to entry, that is, it invites the sort of competition which automatically prevents the continuation of such excessive rates.

The stand-alone cost test does not apply, and cannot be made to apply without disastrous consequences, if railroads are denied the freedom to abandon unremunerative facilities or services. Where that is prevented, a railroad cannot earn adequate revenues if it is constrained by stand-alone cost ceilings on rates in the potentially remunerative portions of its activities. For this reason, it is unwise for public policy to limit the freedom of railroads to curtail unremunerative services without simultaneous provision of public funds to help defray the costs of those services.

The stand-alone cost ensures the equitable treatment of all of a railroad's shippers. By requiring each service or group of services supplied by a rail carrier to contribute revenues less than stand-alone costs, the test assures each shipper and each group of shippers a share in the benefits derived from simultaneity of production--from economies of scope which imply that the total cost to a carrier of supplying many services simultaneously is less than the sum of the costs of supplying them each in isolation from one another. Thus, each shipper is guaranteed some benefit from the revenue collected by the carrier from others. The stand-alone cost offers assurance to each shipper that it will be better off with the existing rates than it would be if it had to fend for itself, as would be the case in the long run if the rail carrier were denied adequate rates.

If the price paid by a shipper is no greater than the stand-alone cost of serving him, then that price cannot possibly contribute to the cost of any facility from which he derives no benefit. This must be true because the stand-alone cost of any facility used by a shipper includes only the (replacement) cost of those facilities after subtraction of any contributions made by any other railroad customers toward the cost of these services. Thus, together, all the customers who share the use of some facilities will provide revenue contributions which do not exceed the costs of the facilities which they use--there will simply be no excess that can be used to defray the cost of facilities unused by any member of this group. The stand-alone cost test, therefore, precludes cross-subsidies among the railroad's different customer groups.

The absence of cross-subsidy under the stand-alone cost test is an appropriate and accepted criterion of equity in the treatment of shippers. Cross-subsidies are properly of public policy concern because they generally lead to a misallocation of resources by encouraging inefficient investment. For the shippers, they may be of concern because they are perceived as unfair. Two groups of shippers may be taken to be treated inequitably if the payments of one of these groups helps to make up for shortfalls in payments by the other. Yet, while shippers who pay more for their service may feel that they are being forced to "cross-subsidize" the other shippers, mere payment of a relatively higher rate is not evidence of a cross-subsidy where fixed and common costs must be covered. Rather, a cross-subsidy can only occur

in an economic sense where a shipper (or a group of shippers) pays more than the total cost of serving it alone. Where no shipper pays more than that amount, differences in their rates simply reflect differing contributions to the common costs of the system, and not cross-subsidies.

Imposing stand-alone cost as a rate ceiling is a form of incentive regulation that avoids introducing distortionary incentives to the railroad with respect to its operations and costing decisions. Since the stand-alone cost is the cost of service by a hypothetical entrant who offers alternatives to the shippers at issue, it is not determined by any of the costs actually incurred by the actual regulated railroad.[7] Consequently, under the system of stand-alone cost rate ceilings, a railroad has no incentive to pad or otherwise increase its expenditures for the purpose of relaxing a regulatory constraint. Further, since the ceilings apply only to services over which the railroad has monopoly power, they do not interfere with the railroad's incentives aggressively to pursue additional traffic and other new business opportunities. Finally, while stand-alone costs may be calculated on the basis of detailed engineering studies and judgements, it is significant to note that they are consistent with the "price-caps" that are becoming so popular today inasmuch as they can be periodically updated on the basis of net measures of inflation and changes in productivity.

2.11 Efficient Pricing Does Not Require Regulatory Control Over the Entire Rate Structure

For prices to be efficient, they must reflect implicitly all of the interdependencies which characterize a rail network. This could be taken to imply that to institute efficient prices for one segment of a railroad's activities (which requires regulatory oversight), it would be necessary to simultaneously regulate the prices for all of the railroad's services. Convincing evidence that such a conclusion is unfounded is provided by the workings of the free market in unregulated industries. In such industries, there exists no authority which coordinates pricing decisions, and yet compatible and efficient prices emerge, their consistency assured by the forces of competition. This is precisely why free and unplanned markets perform so effectively in comparison with those operated by central planners, despite the latter's alleged ability to take interdependencies into account.

It is for this reason that no regulatory control need be exercised over rates of competitive services. Here efficient prices are automatically imposed by the market, and regulatory intervention can only impede the efficiency of the process of rate determination and resource allocation. Also, relatively little control need be exercised over rates set by a carrier whose revenues are still short of adequacy. If total revenue

---

[7] This important property of stand-alone cost is not significantly undermined by the practice of determining stand-alone cost in a fashion that is informed by the operations of the actual railroad. While these operations may provide guidance or even a model for the operations of the stand-alone railroad, the stand-alone cost need not reflect the same decisions as those made by the incumbent, especially if they lead to unnecessarily high costs.

is not yet adequate, the best rates in terms of the public interest in the long run are those that maximize the railroads' net revenues, i.e., Ramsey prices. Any railroad with inadequate revenues has powerful incentives to select such rates. In such a case, the railroad as a whole possesses no monopoly power which offers it excessive profits, and for individual services for which competition is inadequate, the stand-alone test cost provides the requisite protection to shippers. Under these conditions, there is no possibility of unfair competition through cross-subsidy, with noncompetitive rates increased in order to permit noncompensatory prices in competitive markets. For where the railroad's overall revenues are inadequate, any internally subsidized service must be self-destructive--a drain on the railroad's already insufficient revenues. Thus, where overall revenues are inadequate, only the stand-alone cost test need ever be employed in the regulatory oversight of rate setting.

The only case in which more than this minimal regulatory scrutiny may conceivably be required is that in which a railroad is in a position to earn revenues which are more than adequate. Here, there is at least the hypothetical possibility that high prices for one service will be traded off for price reductions in another. Consequently, it may be desirable to devote regulatory attention to prices for services sold on markets from which competition, direct or indirect, actual or potential, is absent. Yet, even here, there are incentives for the railroads to select the efficient Ramsey prices. That is, the interests of the railroads are still likely to be served best by the prices which best serve the public interest--though it must be admitted that the incentives for it to do so are apt to be somewhat less powerful than those in the prevailing case of inadequacy of revenues.

There is one principal source of incentives for a carrier capable of earning adequate revenues to adopt efficient pricing, even though its net revenues are constrained by regulation just to cover its capital costs and no more. Such a rail carrier is motivated, perhaps more than other firms in similar circumstances, to maintain its traffic base and to guard against substantial diversion of its traffic to suppliers already in operation or to potential competitors. This is because a large portion of a rail carrier's capital stock is nonfungible, or sunk, so that significant losses of traffic would cause losses of revenue far greater than the costs that would thereby be saved. Consequently, a rail carrier with adequate revenues has a particularly compelling incentive to set rates in a manner that will discourage defections of shippers and market erosion to competing suppliers of transportation services, both in the short and in the long run. It may be clear intuitively that among the pricing policies that generate adequate revenues, it is Ramsey pricing that most effectively discourages such defections and market erosion. This is simply because at any one time the Ramsey prices yield shippers the greatest total net benefits possible from prices which yield adequate revenues, and therefore offer shippers the smallest feasible inducement to divert their traffic.

In sum, regulation need not take on the overwhelming task of control of all of a railroad's rates, simply to assure an appropriate choice of prices in those circumscribed arenas requiring regulatory attention. Elsewhere, the forces of competition and the self interest of railroads constitute powerful mechanisms which can do the job efficiently and automatically, using the crucial demand information possessed by the railroads, which is certain to be more complete and more accurate than any demand data a regulatory agency can hope to assemble.

2.12 Contestability and the Scope and Structure of Regulation

Contestability is an apt benchmark for the railroad industry, while the familiar perfect competition benchmark is neither attainable nor desirable for the railroad industry where economies of scale and scope are substantial. In this industry, attempts to approximate perfect competition may in fact be highly inefficient and contrary to the public interest. In any case, the theory of contestable markets demonstrates quite clearly that neither large size nor fewness of firms necessarily means that markets need function unsatisfactorily. Indeed, a variety of market forms far removed from perfect competition may perform well for the public interest so long as such markets are structurally contestable. If an industry is contestable, then it is best left on its own devices with no government interference, even if it is composed of a very small number of large firms. Impediments to entry and exit, not concentration or scale of operations alone, are a primary source of interference with the public-interest workings of the invisible hand.

Contestability focuses increased attention upon entry barriers and their defining characteristics. High fixed costs and the consequent economies of scale, for example, have traditionally been considered as impediments to entry; contestability analysis shows, however, that they need not permit excessive profits or prices or any of the other manifestations usually associated with market power. It is the presence of sunk costs rather than economies of scale that is of vital importance for market performance.

The theory of contestability offers an improved set of guidelines for appropriate government intervention in the structure and conduct of firms and industries, that is, of the rules to be followed by the regulators in those cases in which their intervention is called for. In addition, it provides economically sound criteria distinguishing between those cases in which intervention by the public sector is warranted and those in which it is not. The theory of contestability is the framework from which was derived the following precepts for railroad regulation already discussed above:

(i) Permit a private sector railroad to have freedom of pricing and operations on services that face effective competition in the relevant market, whether from other railroads, other transportation modes, other origins, other destinations, or other commodities.

(ii) Permit a railroad to set prices that are responsive to differences in demands, as well as to differences in marginal costs, and further to enter into voluntary contracts with shippers that have individualized terms, conditions, commitments and/or compensation mechanisms.

(iii) Constrain the prices that a railroad sets to "captive shippers," over whom the railroad has monopoly power, by the stand-alone costs of the shipper's service (or by a comparison of revenues and stand-alone costs associated with any larger group of shippers' services), and by the stipulation that the railroad's prices do not generate earnings that persistently exceed the railroad's replacement costs, including a competitive return on capital.

In addition, contestability is a fruitful framework for analysis of issues pertaining to the vertical structure of an industry. For one thing, in a perfectly contestable market, survival against potential competition requires that a firm undertake efficient vertical relationships and structure itself efficiently along vertical as well as horizontal and conglomerate dimensions. For another thing, contestability theory suggests consideration of the idea of separating firms vertically in order to segregate portions that need regulation from those that do not because of their degrees of competition or contestability.[8]

This idea emerges from the application of contestability theory to regulatory policy where sunk costs are not pervasive in an industry, but rather are centered in a particular sector of its operations, such as the track, way and structures in railroading. By isolating the activities with which the heavy sunk costs are associated, their need for regulation can be quarantined. By placing relations with the remainder of the industry at arm's length, to the extent that is permitted by economies of scope, it may be possible to leave the operations of the bulk of the industry safely to the free market, drawing a regulatory net over only the segment of the activities of the industry that are inextricably associated with heavy sunk costs. Thus, contestability suggests a flexible case-by-case regulatory approach, whereby activities subject to effective competitive pressures from the active or potential supply of substitute services and markets in which efficient technology does not require significant sunk costs are freed from traditional regulatory constraints and are permitted open entry and more flexible pricing.

2.13 The Vertical Structure of the Railroad Industry

The historical model of railway operations is the monolithic organization, whereby a single entity controls all facilities, operating and administrative functions, and determines what services to provide to significantly captive markets. The conditions that generated this model no longer exist in most countries, and governments have had to consider fundamental restructuring of the railway entity itself, and the

[8] See Bailey (1981) for the first expression of this idea.

relationship between the railway and the State. The objectives for such restructuring have properly included injection of more innovative and efficient management, reduction of railway deficits and burdens of public subsidies, increased competition with other transport modes, and more responsiveness to the needs of emergent capitalist enterprises. Five generic options can be identified for the vertical structuring of railways, addressing the set of relationships between the railway entity and other transportation entities (both rail and otherwise), the markets served and the functions performed. These functions include ownership; improvement and maintenance of the fixed facilities; control of operations such as dispatching and freight classification; train movement; equipment provision and maintenance; marketing; and financial control and accountability.

3. THE GENERIC OPTIONS FOR VERTICAL RAILWAY STRUCTURING[9]

Option 1: The Monolithic Railway. The frequent status quo is the traditional monolithic approach, under which the railway is an integrated entity owning and operating its own facilities and vehicles. Typically, the monolithic entity lacks financial incentives and desegregated information on profitability, is (at best) production-oriented, is unresponsive to market demands for services, and is hierarchical (if not bloated) in organizational architecture.

Option 2: Lines of Business Organization. Railway entities may be reorganized and accorded financial responsibility for lines of business to foster comprehensive business planning, market-sensitive and cost-sensitive decisions, and greater responsiveness to demand for various services. British Rail, for example, has divided itself into five lines of business that are financially accountable to top management and that "purchase" service by contract from an operating department that is organized along a matrix of regional and functional lines. By so doing, British Rail hopes to give commercial sectors a profitability objective, and to give noncommercial lines of business incentives to reduce their losses.

Option 3: Competitive Access. Competing railway companies would have exclusive control over some trackage, but would also have (and give) the right of competitive access over the trackage of (to) other companies. Some forms of competitive access include joint terminal agreements and/or conferrals of trackage rights, whereby one railway obtains the right to use freight handling facilities and/or the line haul tracks of another railway at a particular location or along a particular route. A further characteristic of this option would be arrangements for interlining traffic that is handed off between distinct railroad entities, in their preference sometimes to utilization of trackage rights. In the U.S., railroads do a great deal of interlining, under terms that are largely unregulated, perform reciprocal switching under terms that

---

[9] For an illuminating analysis of the options for restructuring railroads see Moyer and Thompson (1992).

are subject to regulation, and exercise trackage rights that are sometimes freely negotiated and sometimes result from regulatory mandates (that were mostly put into place in the context of settlements of disputes over rail mergers).

Option 4: The "Wholesaler". Under this option, the railway entity could own and operate the fixed facility and perform all operations on behalf of marketing entities which would be the "retailers". This would mean that the railway itself would only haul trains, but would do no marketing to shippers. In Australia, for example, freight forwarders function as retailers using the state railways' "wholesale" services. These forwarders provide multimodal transport, and conduct a deregulated trucking business. They control their own rail terminal and yard operations and negotiate on the open market with the railways to charter unit trains with agreed-upon service specifications. This permits competition among efficient intermodal "retailers" to flourish, despite a state or private monopoly on railway ownership.

Option 5: The "Toll Rail" Enterprise. Under this option, the entire fixed facility, except for exclusive facilities, would be the property and responsibility of one owner. There could be one or more authorized user(s), which would pay tolls for use of the facility. This approach differs from "competitive access" in the following respects: under the "toll rail" approach, separate entities provide the fixed facility and conduct operations, whereas under "competitive access", more than one entity operates in a given market over a particular fixed facility. Since 1988, Sweden has implemented a separation of fixed facility from operating functions. The U.K. has recently moved in this direction by establishing a separate entity to hold and manage the rail system's assets associated with the track and road bed. And the European Union has articulated the policy principle that its members move in the direction of separating rail operations from the fixed facilities.

3.1 Analysis of Monolithic Railway Option

The monolithic railway option is largely a strawman from today's perspective--i.e. no one would deliberately choose it for the public interest. Nevertheless, it is an option that has in fact been chosen all too often, either for private interests in monopoly control, or more often, for the political benefits that could be collected and disbursed through a state-owned monolithic railway. It is predictable that a state-owned railway enterprise would fail to be beneficially responsive to the needs of shippers, and would instead be politically responsive, at the expense of efficiency of operations and of stimulus to the economy.[10] It is equally predictable that a privately owned railway that were exposed to excessively controlling and economically arbitrary regulation would also be without incentives for efficiency and for market responsiveness. Either way, financial deficits would be a natural consequence, as the railway entity

[10] See Willig (1994).

failed to succeed in attracting traffic from alternative modes and geography, as it expended inefficiently on costs, and as it allowed its facilities to suffer from deferred maintenance and replacement.

3.2 The Need for Restructuring

There is no doubt but that sustained economic growth and prosperity require transportation that is responsive to shipper needs and demands, as well as to marketplace opportunities for innovation. It is clear today, too, that a railroad organized and controlled according to the monolithic model must be restructured in order to contribute best to the economy and to avoid being a significant impediment to growth and prosperity.

One key element of restructuring is to develop internal organization of rail entities that provide managerial incentives, information, and decision-making decentralization that conduce to efficiency, market responsiveness, and fiscal responsibility. Thus, Option 2 is certainly crucial for restructuring, whatever else is also entailed. It should be recognized that an internally restructured railroad enterprise may show lower technical operating efficiency by some traditional measures (e.g., coach-kilometers per locomotive-kilometer), but may achieve greater responsiveness of each service to customers' needs and willingness to pay. Economic productivity and the customer's interests are best promoted by minimum total logistics costs, not just the lowest railway rates accompanied by minimum service quality.

Another key element of restructuring is to unleash market forces of competition, to the fullest extent that is consistent with opportunities and other elements of efficiency. It is difficult to predict what are efficient and market responsive vertical relationships and combinations of logistical roles among various rail entities, truckers, barge operators, port operators, warehouses, forwarders, etc. The U.S. experience confirms what theory predicts: decentralized market-oriented decision-making that is freed from excessive regulatory control and that is energized by market incentives is the surest means of finding and implementing efficient and innovative solutions to problems posed by transportation needs.[11]

Options 3, 4, and 5, as defined above, are approaches to restructuring that have the potential for bringing more competition and more market decision-making into the domain of railroading and its vertical relations. Which of these options is the best choice is a complex policy decision with many important dimensions to consider. Below we offer analyses of the options that may help to clarify some of the important considerations.

3.3 Analysis of Structural Separation -- Options 4 and 5

The options that separate ownership of facilities from other rail functions such as train operations and marketing have generated much attention of late, and deserve serious analysis. These options have

---

[11] For example, see Willig and Baumol, op. cit.

considerable appeal because they seem to mitigate the difficult problems blocking comprehensive rail deregulation that are associated with the roadbed costs that are largely sunk. Fixed costs are large because of the infrastructure (track, stations, etc.) that must be provided before any trains can run on a route. Duplication of infrastructure would generally be inefficient, and natural monopoly cost conditions therefore characterize physical network provision. These fixed infrastructure costs are largely sunk because the assets are of minimal value for other purposes. For example, embankments and cuttings, the rail formation and the platforms are fixed *in situ*--they are sunk, committed irreversibly to a specified market. The sunk nature of infrastructure costs creates significant entry barriers, especially where there are natural monopoly conditions as well.

The cost conditions relating to the operation of services on the physical network, on the other hand, may be more consistent with active and potential competition. To operate a service it is necessary (at least) to have trains, staff, support, and rights of way. Although there are inevitably some sunk costs in hiring staff and buying or leasing rolling stock, they are small in relation to the massive sunk costs of establishing network infrastructure. Locomotives and freight cars constitute capital on wheels, and most of their cost might be easily and quickly recovered by rolling them to other markets.

Thus, it is possible that if ownership of track and trains were separated, with the track assets held by the government, by a consortium of the operators, or by a regulated private entity, then there could be vigorous active and potential competition over railway services provided by operators with equal access to the utilization of the roadbed. Consequently, these operators need not be subject to regulation, and they would have all the powerful incentives that accompany competition to be efficient and responsive to the needs of shippers and a growing entrepreneurial economy.

However, there are several links in this chain of policy reasoning that may be inapplicable or wrong in a given set of realistic circumstances.

(i) The provision of many innovative and market-responsive rail services may require specific investment in infrastructure, such as maintenance or upgrading of way and structure facilities, construction of loading and transhipment facilities and building spurs of track to reach a shipper's location. It may be difficult and inefficient for any operator (or retailer) to coordinate, as necessary, with the infrastructure monopoly (or wholesaler) entity, especially if their incentives with respect to investment behavior are not in harmony. The investment incentives of the infrastructure monopolist (or wholesaler) will, of course, depend critically on whether it is a state owned entity, or, if it is in the private sector, on the character of its regulation.

(ii) Efficient, safe, and delay-minimizing utilization of track and yard facilities by trains, cars, and shipments requires close coordination in accordance with priorities that are driven by considerations of both operations and shipper sensitivities. Competing operators (or retailers) will compete vigorously and acrimoniously over scarce or congested infrastructure facilities (or wholesaler services), and constantly sorting out their claims will be important for the overall efficient and responsive operation of the rail system. This would be difficult for an unintegrated system with a monopoly infrastructure entity, but it seems virtually impossible to accomplish efficiently under conditions of rules against discrimination and infrastructure (or wholesale service) pricing that is either tightly regulated and/or, for a state enterprise, politicized.

(iii) It is plausible that the freight hauling operations on all or part of the rail system in question comprise a natural monopoly, even disintegrated from the infrastructure. The economies of scale and scope that arise from running long trains, from blocking many different shippers' freight in classification yards, and from efficient utilization of yard facilities, crew, and rolling stock, all are associated with operations, rather than with infrastructure. Consequently, a separated operations firm may be a monopoly, and it may have considerable market power unless potential competition is a powerful force.

(iv) For potential competition to be powerful, an entering operator must perceive that significant sunk investments in rolling stock and in specialized facilities can be avoided. Locomotives and freight cars may indeed be an example of "capital on wheels" so long as they can be transported to alternative points of gainful utilization without substantial costs. While this is likely to be the case for services provided in the middle of a landmass with a rich rail network ready to accommodate the cars, it may not be the case for more specialized cars or for a more isolated market. Also, the entering operator may not have yard, loading, car maintenance, or spur facilities available unless new and significant sunk investments are made. For these to be available on equal terms with the incumbent operator, it must be the case that the infrastructure entity made the needed investment as part of its role in the system. But the more of the entrepreneurship and risk-taking investment that must be accomplished by the infrastructure entity (or the wholesaler, under that option), the less it is that is gained by the separation, since the infrastructure (or wholesaler) entity is either a state-owned or a tightly-regulated private sector monopoly.

(v) Efficient pricing to cover replacement costs is made more difficult by separation. Where economies of scale are important, efficient pricing to cover replacement costs requires that shipments of different commodities on different origin-destination routes bear prices with different relationships to marginal costs. If it is the case that the operator (or retailer) firms can readily evade price discrimination on the part of the infrastructure entity (or

wholesaler), so that different prices cannot be collected by the infrastructure entity (or wholesaler) for facility utilization (or for wholesale service utilization) by different shippers of different commodities, then it will be difficult if not impossible for the costs of the infrastructure to be defrayed by Ramsey prices. At the extreme, a regulated infrastructure (or wholesaler) entity charging competitive operators (or retailers) an equal price for each ton or each ton-mile of freight that utilizes each of its facilities is, in essence, recreating a system where prices are set according to fully allocated costs. As discussed above, such pricing can be a prescription for inefficiency and financial disaster.

Thus, it is clear that separation of operations from infrastructure in a railroad system is no panacea for regulatory problems. Instead, as a policy direction, it must be compared with the leading alternative.

3.4 Analysis of the Competitive Access Option

This option is most clearly distinguished from the separation option just discussed by the allowing of integrated operations by the rail entity. It is superficially easy, albeit mistaken, to identify an integrated carrier with the case of the monolithic carrier, because it is tempting to jump to the conclusion that an integrated carrier would make it very difficult for other entities to participate in its business. This option contemplates a requirement that the integrated carrier make its facilities available to other entities on a "fair and equal basis." However, if the integrated carrier has strong incentives to keep other entities out, it is unclear how effective such "equal access" mandates are likely to be. The rail industry in the U.S., like other regulated industries in the U.S. as well (e.g. gas pipelines, telecommunications, and electric power), has seen many disputes with claims of "unfair" and "unreasonable" exclusion from a carrier's facilities, despite rules of "equal access."

Thus, it is key to an assessment of this option to analyze the incentives of the integrated carrier to accommodate others wishing to participate, and able to participate efficiently, in the provision of service.

It is clear that if the integrated carrier is regulated in a fashion that permits higher prices to be charged to captive shippers if the carrier does more of the business, than the carrier would have incentives to exclude other participants.[12] Likewise, if the integrated carrier is constrained by regulation in the amount it can earn from the portion of service it provides, when it does cooperate with another entity, then it has incentives to undermine or avoid efficient cooperation in order to enlarge the portion of service it

[12] One example of this effect arises under rate-base rate of return regulation, as was understood by Averch and Johnson in their seminal paper.

provides.[13] In addition, the integrated carrier would be motivated to exclude an efficient participant if by so doing the carrier would weaken, in a predatory manner, the competitive impact of that entity in another market. Under classic rate-of-return regulation, or under a system of regulated "divisions" which specifies what an integrated carrier can earn from a cooperative movement, both features of U.S. rail regulation at one time, an integrated carrier does have incentives to undermine efficient cooperation.

In sharp contrast, under the regulatory system that has been described above as well-serving the public interest, an integrated carrier would generally have a real profit motive to cooperate with an efficient participant in its business. Here, it is not "divisions" that are specified by regulation, even on service provided to a captive shipper. Instead, the described stand-alone cost rate ceiling applies to the price charged to the shipper, and cooperation with an efficient entity enlarges the pot of returns available from the service, enabling more money rather than less to be earned by the integrated carrier. Consequently, except for the rare possibility of predation, an integrated carrier would have ordinary business incentives to find and to cooperate with efficient participants in its business, and to negotiate with them terms that would be mutually beneficial. This is just a railroad version of business "make-or-buy" decisions in other industries.

Despite the prevalence of efficient incentives on the part of integrated carriers under the form of regulation described here, it is useful and wise to augment the system of regulation with a fallback set of standards to apply should disputes about predation through competitive access arise. In short, an integrated carrier that possesses a "bottleneck," i.e. a facility without which the complainant cannot reasonably offer its services to the shipper, should not exclude the complainant by refusing an agreement that would be fully compensatory of all its costs, including opportunity costs.[14] For example, if another carrier, or an operator, sought to participate in a freight movement that represented new business for the integrated carrier, then it is to be expected that the latter would negotiate in good faith and not exclude the other entity if an agreement could be found that would at least cover the incremental costs of the integrated carrier. If another carrier sought to handle some freight part of the way that the integrated carrier would otherwise handle itself, then it is to be expected that the integrated carrier would accept an agreement that earned it a larger net contribution of revenues above incremental costs than it would earn if it handled the freight without the other participant. Here, the contribution that the integrated carrier would earn on its own is part of the opportunity costs it faces from cooperating with the other participant. These same principles apply to interlining, trackage rights, car hire, or any other form of cooperation or participation through the employment of a bottleneck.

---

[13] For a more complete discussion of these cases, see Ordover, Sykes, and Willig (1985).

[14] This standard was first developed in Ordover and Willig (1981).

"Efficient component pricing," or "parity pricing," is the name that has been given to the principle that an integrated carrier should offer the services of its bottleneck at a price that yields it the same contribution that it would earn from performing the end-user's service itself.[15] Behavior consistent with this pricing of bottleneck services, or more generally with the anti-predation rule just articulated, leads to efficient vertical relations, and is thereby consistent with non-predatory incentives under the regulatory system we have described. Such pricing of bottleneck facilities does not place additional competitive pressure on pricing to shippers, since it is based on the contribution that could be earned from the shipper's service at the extant shipper's price. However, it does generate incentives for efficient combinations of transport services to make it to the market, it does provide quality and cost competition among potential and actual participants for the role of being part of the efficient combination, and it does help to assure that those with efficient innovations in logistics or in marketing of transport services will be able to work with carriers to implement their ideas.

3.5 Comparing Separation with Competitive Access

The primary virtue of separation as a policy option is that it may permit active or potential competition to reign among rail operators or retailers--with corresponding assurance of efficient selection among them for provision of their services at efficient prices. At best, separation will accomplish this end, but leave unresolved the difficulties with regulation of the provision of the services of the infrastructure, or bottleneck, assets of the railroad network. Prices charged to shippers will be at least the sum of the competitive prices for the services of the operators (or retailers) and the regulated prices for the services of the infrastructure entity (or wholesaler). They are unlikely to be fully Ramsey efficient prices for the coverage of replacement costs, because of difficulties of reflecting shippers' differences in demands in the prices charged for infrastructure services. At the same time, separation may create serious coordination problems, loss of economies of scope, and otherwise unnecessary transactions costs. In addition, rail operators may not face effective active and potential competition, undermining the potential for realizing the primary benefit of the option.

In comparison, the competitive access option could also be fraught with problems, when incentives of bottleneck holders are adverse to efficiency and competition. A variety of solutions to competitive access problems have arisen in industries seeking to replace regulation with competition. Typical examples include: mandatory interconnections with competitors and line-of-business restrictions in the telecommunications industry; "unbundling" of the transportation and energy components of price in natural gas markets; and equal access to marketing channels (e.g., computer reservations systems) in the airline industry. In designing rules that govern vertical relationships among competitors formerly subject to

---

[15] See Baumol and Sidak (1994).

economic controls, regulators need to address a common basic problem--how to assure that pricing and terms of access by "nonintegrated competitors" to the restricted portions of the network will be implemented so that competition on the merits will work to assure that the efficient alternatives do successfully participate in the provision of end-users' services. The compensation for and terms of access should not distort the process by which prices are adapted to consumer preferences and demands for transportation service. Prices should be sufficiently high to be compensatory to the "landlord" railroad, yet not so high as to preclude efficient operations by the "tenant" railroad. Where incentives are significantly adverse to these goals, experience teaches that rules are too easily evaded, and disputes seemingly never-ending.

It is thus fortunate that under rail regulation that focuses on the levels of rates charged to shippers, rather than on other prices, such as those charged for access to bottleneck services, incentives are generally for the promotion of efficient vertical relationships. As a result, if integration is permitted under this system of price regulation, then the outcomes are predictably consistent with efficient participation by the integrated carrier and with other nonintegrated carriers as well, on terms that permit compensatory support for the efficient participants. Further, prices to shippers can be selected in accordance with Ramsey efficiency, even as they are constrained by regulation where the carrier has monopoly power. Moreover, unlike the virtues of separation, the efficiency of the outcomes of competitive access does not depend on the absence of economies of scope, on the absence of coordination problems without integration, and on the competitiveness or contestability of rail operations.

On the other hand, separation of track assets from operations is likely to be a particularly attractive option where a dense and extensive rail network permits many operators to function, and to provide both active and potential competition to each other. Another favorable factor is a mature and well-developed set of fixed facilities, so that there is relatively little extent to the domain of new infrastructure investments where incentive problems are more likely to arise. Where this factor does not apply, it will be important for regulation of the infrastructure entity to permit it to enter into medium or long term contracts with shippers or with operators that themselves have contracts with shippers, so that the risks and rewards from investments can be efficiently shared by shippers, operators, and the infrastructure entity. The impediments to Ramsey pricing that separation might cause would be rendered insignificant to the extent that the infrastructure entity does not attempt to recover its sunk capital costs from "tolls" levied on traffic. If the infrastructure entity is expected to seek recovery of its replacement costs, then it should be permitted and even encouraged to implement forms of price discrimination that help to bring shippers' prices in line with principles of Ramsey efficiency. Finally, there may well be circumstances where a monolithic railway system cannot be converted to one with functioning competitive access because of imbedded business culture and entrenched management. Here, the act of separation is so revolutionary that it may unsettle the business culture in a productive fashion, and force reassignments of management that permit

implementation of the necessary internal reorganizations of responsibilities, roles, incentives, and information flows.

4. CONCLUDING REMARKS

We have outlined a set of principles that together add up to a program for restructuring the relationships between government and railroad entities. These principles point towards a great deal of reliance on market forces to shape prices and logistics of services. At the same time, the principles include economically appropriate protections for any captive shippers and for any carriers that may be excluded or foreclosed from participation for anticompetitive reasons. On the subject of restructuring, we have pointed out that internal managerial reforms are necessary, as are policies that address railway vertical relationships. The two leading candidates, separation of track from operations in different business entities, and incentives and fallback rules for competitive access, were compared on several dimensions, and their relative levels of appeal were found to depend on a variety of characteristics of the business environment.

It can be expected that restructuring along the lines we have suggested here, providing a greater emphasis on marketing effectiveness, will result in a more profitable railway that can better hope to cover its costs, in the case of commercial services. Needed noncommercial services should be carried out on the basis of an explicit agreement between the railway and government that views public service obligations as a business relationship between a customer (government) and the contract supplier (railway). This would help to ensure that noncommercial services are more effective in fulfilling public policy objectives, while eliminating an insuperable drain on revenues that would condemn the railroad to insufficient investment, and eliminating cross-subsidies that make it difficult for the railroad to maintain its efficient competitiveness against other modes.

Appendix A. TECHNOLOGY AND THE STRUCTURE OF RAILROAD COSTS

The output of the rail industry is multidimensional by its very nature. Railroad firms produce different types of transportation services for different users at different origins and destinations at different times and at different levels of quality. Consequently, the mix of output and shipment characteristics can have a major impact upon the costs of any given firm. For example, railroads specializing in coal traffic have very different cost characteristics than those specializing in movements of general manufactured commodities.

The most striking feature of the railroads' cost structure is the high incidence of costs that cannot properly be attributed to any particular service at a particular point in time. That is, a significant portion of costs are incurred on behalf of several activities and do not vary with the amount provided of the service in question. These unattributable expenditures reflect both joint and common costs. "Common costs" are costs shared by two or more services in variable proportion. For example, a terminal represents a common cost; it is used by different services in varying proportions. More generally, the outlay on track and way and structures between points A and B is a common cost for all movements of whatever commodities are shipped between A and B over that route. "Joint costs", on the other hand, are costs shared by two (or more) services in fixed proportions. A backhaul movement is the classic railroad example.

The structure of railroad costs has important implications for the competitive structure of rail markets. It is sometimes mistakenly inferred from statistical evidence of constant returns to firm size that a competitive equilibrium with marginal cost prices covering total costs would be sustainable in the rail industry. Such reasoning neglects the critical fact that indivisibilities in rail technology make increasing returns to scale in total costs endemic, and small numbers competition inevitable. A rail link between two points requires lumpy investment in way and structures with associated highly significant economies of traffic density--unit costs fall with output, letting all factors of production vary on a given route or route structure.

A.1 Fixed and Variable Costs

A fixed cost is one that is necessary to provide a service or group of services, but whose magnitude does not vary with changes in the quantity of a service that is planned to be or that is in fact provided. For example, if a railroad is to run between A and B, there is a minimum outlay on track and roadbed that must be incurred, even if the trains run virtually empty. The service can be discontinued altogether; but even in the longest of long runs, its roadbed cost cannot be reduced to a negligible level if the amount of the service is to be positive. Also, a loading facility may be necessary to transport coal

efficiently between points A and B, but its cost may be unchanged if the amount of coal transported is doubled or halved. Common costs are often fixed (e.g., the basic portion of the outlays on track and way and structures between A and B may be both fixed and common costs).

Fixed and common costs are quite different from variable costs. Economists employ two fundamental cost concepts in defining variable costs--marginal cost and incremental cost.

The marginal cost of a service is the additional cost that would be incurred in order to supply an additional unit, or the saving in total cost that would be made possible by supplying one less unit. As such, the marginal cost of a rail service is the per-unit opportunity cost to the rail carrier of the level of a service's volume. The term "opportunity cost" refers to the value a resource can contribute if it is used in some alternative occupation instead of the one to which it is currently assigned by the railroad. Thus, marginal cost is similar in meaning to unit incremental cost and to the true economic variable cost. However, the definition of marginal cost makes it clear that it should include the traffic-sensitive costs of capital facilities that are fungible and economically attributable to the service, as well as the more obvious cost components such as fuel, labor, and traffic-sensitive maintenance and replacement costs.

For example, locomotives and other rolling stock employed for some period of time to provide a given rail service have a significant opportunity cost for a rail carrier. If not utilized to supply the service in question, they could instead be gainfully utilized elsewhere in the rail network, by the rail carrier at issue, or by some other carrier. Assuming that at least some carriers do not have excess supplies of the equipment in question, or their functional equivalents, a decrease in the quantity supplied of the service would release equipment that could decrease or delay the need of some carrier to lease or purchase stock for replacement or expansion. Consequently, it follows that the opportunity cost of the rolling stock is its replacement cost, at the current cost of capital. Thus, the marginal cost of a given service includes the costs of fungible capital goods that are utilized, such as locomotives and other rolling stock, at the current cost of capital for the period of time during which they were so employed. Of course, the marginal cost of a service also includes the wear-and-tear on capital assets and the required maintenance expenses that the supply of the service causes.[16] However, the costs of facilities that are fixed or common are not included in the service's marginal costs.

The incremental cost of a service is the cost per unit of service necessary to provide the entire service, or the cost avoided by not providing the service, given all the other services supplied. The term

---

[16] For example, it is clear that the passage of rail traffic causes wear and tear on track, ties, and ballast, and that this in turn shortens the lives of the assets. Consequently, one element of the marginal cost of rail traffic arises from the hastening of the time that the assets it utilizes must be replaced--to wit the present discounted value of the capital cost of the assets' value, over the time period that their needed replacements are advanced.

"avoidable cost" is also used to describe the cost per unit of service that could be avoided by not providing a particular service.

The important conceptual point here is that a railroad's total costs are composed of some costs which vary with the amount of a particular service provided and others that do not. This is obvious enough, but considerable confusion is often engendered when the additional point is made that in the long run virtually all fixed and common costs can be varied. The reason is simply that in the long run virtually all assets must be renewed or replaced. At the date when the decision regarding renewal or replacement of the fixed factors of production required to supply a service or group of services is under consideration, the costs involved are incremental to that service or group of services. If it were decided no longer to provide those services, the costs would not be incurred.

This obviously does not mean that there is no economic distinction between variable costs and fixed and common costs. What it does mean is that the perspective of the decision maker is very important. When a railroad is making decisions regarding the incremental costs of adding a particular service (or the avoidable costs of eliminating a service) given existing capacity, the short-run variable costs of service will include only the additional costs of production imposed by that service. Rarely will this include the full measure of long-run fixed costs. In contrast, when a railroad is making the long-term decision whether it is in its economic interest to replace a portion of its rail network (or to make an entirely new addition to its network), the "long-run variable costs" of the service or services the railroad plans to offer will include all the fixed costs which will become sunk (i.e., irreversible for a significant period of time) once they are incurred.

A.2 Sunk Costs

Long-run fixed costs are those costs that are not reduced, even in the long run, by decreases in output so long as production is not discontinued altogether.[17] But they can be eliminated in the long run by total cessation of production. Sunk costs, on the other hand, are costs that (in some short or intermediate run) cannot be eliminated, even by total cessation of production. As such, once committed, sunk costs are no longer a portion of the opportunity cost of production.

Sunk costs need not be fixed and, even more important, fixed costs need not be sunk. To operate with current production techniques, a railroad requires at least a locomotive and one car, the costs of which must be included among its fixed costs. Yet, because they constitute capital on wheels, most of their cost can easily and quickly be recovered by rolling them to another market, should the railroad's

---

[17] See Baumol and Willig (1981), and Baumol, Panzar, and Willig (1988).

management decide (and be permitted) to close down the line in question. Thus, little or none of this portion of fixed cost is sunk, in contradistinction to the roadbed cost, which typically is sunk. While bridges, ballast, rails, and ties can also be moved from one route to another, they can be moved only at considerable expense.

The distinction between sunk and fixed cost is not a mere technological quibble. It makes a substantial difference for the design of appropriate public policy if the costs of the firms in an industry include the one rather than the other. Sunk costs contribute to entry barriers which, as is well known, can give rise to monopoly profits, resource misallocation, and inefficiencies. On the other hand, fixed costs do not constitute barriers to entry and do not entail the misallocation problems to which entry barriers lead. Fixed costs are not, and do not raise, entry barriers unless they also happen to be sunk.

A.3 Economies of Scale and Scope

The issue of whether a firm's total costs will be recovered from prices that are equal to the firm's marginal costs of supply is logically equivalent to the question of whether the firm's operations are characterized by economies of scale, or, in alternative terminology, increasing returns to scale.

For multiproduct railroad firms, economies of production could exist due to the level of supply of all the firm's outputs (economies of scale), as well as due to the breadth of the set of services supplied (economies of scope). Economies of scale are exhibited where equiproportionate changes in the levels of all services provided would require a less than proportionate change in the level of efficient costs. In addition to economies deriving from the size or scale of a firm's operations, there is also the possibility that cost savings may result from simultaneous production of several different outputs in a single enterprise, as contrasted with their production in isolation, each by its own specialized firm. That is, there may exist economies resulting from the *scope* of the firm's operations.

There are substantial economies of scale in the provision of some rail services, whether focused on particular routes or types of freight, which result from the heavy fixed costs associated with rail operations. To transport even small amounts of freight, a railroad must generally incur the costs of track, right-of-way, locomotive power, crew, and certain facilities. These costs do not rise proportionately with traffic volume. As more traffic uses a section of a roadway, very few additional fixed costs are incurred, and the extant costs are spread over more traffic. A single track line can handle large amounts of traffic before a full second track must be added or advanced signalling systems installed. Scheduled trains can be made longer to accommodate more cars on the same origin-to-destination route without proportional increases in the costs of locomotive power and crew. The more freight that is scheduled to traverse the same route, the larger can be the preblocked movements, with correspondingly less reclassification yard

activities and time needed, and with more opportunities to run efficient through-train service In short, additions to the levels of rail services supplied do not entail proportionate additions to the levels of expenditure required for fixed plant, for equipment investment, and for operating expenses. This is precisely the hallmark of economies of scale. Fixed costs, of both the sunk and fungible varieties, per ton of freight fall as traffic volume increases, and so cost efficiencies may be associated with single carrier provision of rail services.

Another advantage of firm integration in the rail industry arises from potential economies of length of haul. With fixed terminal expenses, longer hauls normally imply lower costs per mile. In the presence of such economies, a railroad with an integrated nationwide system will sometimes have a cost advantage over competitors that make and accept interline shipments to and from other railroads.

Increased firm size may convey cost advantages because of specialization and massed-reserves economies. A large firm may employ a more richly specialized array of accounting, finance, marketing, engineering, research, and legal talent than a smaller competitor. This may be reflected in lower administrative costs and/or higher productivity. It can mass its cash balance reserves and spread production, market, and financial risks over a larger volume of activity--the diversification of the portfolio of transportation services offered by a large railroad creates an overall system risk factor that could be substantially less than the risk associated with investment in just one of those services. A large railroad firm with an integrated network may also realize economies in equipment investment. In general, railroads attempt to minimize the need for new equipment purchases by using equipment interchangeably throughout the system. When cars and locomotives are needed at some shipping point, the railroad can immediately send them out of the most convenient distribution center. Operations with assigned equipment require more switching than those which draw their equipment from common pools. In addition, the ability to use locomotives interchangeably reduces the number of reserve locomotives needed to protect against equipment failures, repairs and inspection. A larger railroad firm may, therefore, obtain the same degree of protection at lower cost relative to total capacity carrying costs.

Another pertinent feature of the railroad industry is that there are substantial economies of scope which result from the common costs of rail operations. Outlays on rails, ties, rights-of-way, yard facilities, locomotion, and train crews are among the many common costs of rail operations incurred in carrying a variety of types of freight between a variety of origins and destinations. These shared costs confer economies of scope on carriers offering a multiplicity of transportation services: a carrier that provides an array of services can do so at a lower total cost than a set of carriers producing each service separately.

A.4 Economies of Size vs. Economies of Scale and Density

The overall size of a railroad is likely to be quite independent of the amount of traffic that travels on any of its routes--that is, a large firm may have short or long hauls and high or low traffic densities between different points. There has been a serious confusion between economies of scale, economies of size, and economies of density, and a concomitant failure to specify clearly which is being measured.[18] Economies of scale are carefully defined to refer to a long-run average cost curve that declines as the quantity of the firm's output of a given collection of services increases.

Comparing the average costs of railroads that have different sizes of route networks does not provide information directly relevant to economies of scale, because such railroads do not supply different amounts of a given collection of services. Instead, they likely offer quite different collections of services as a result of their different route mileages and architectures. To emphasize the point that the correct and relevant measure in railroading is the degree of scale economies that relates to traffic volume on each particular route, rather than to the extent of the traffic on an overall and possible growing system, these economies have come to be termed economies of density. Thus, the critical determinant in pricing and (dis)investment policies is whether or not there are economies of density. It is therefore important to assess the extent to which unit costs decline as output increases holding the route system, or miles of rail line constant. A small firm with high traffic density could potentially have lower average costs than a large firm with low density.

Economies of density are normally attributed to declining average capital costs. However, the provision of rail service entails more than simply installed capacity; it includes minimal (and often indivisible) amounts of crew, engines, maintenance, etc. Indeed, recent empirical studies indicate that the maintenance of way and structure and transportation expense (mainly fuel and crew wages) account for a significant portion of the estimated economies of density--approximately two-thirds of these economies are due to variations in unit operating costs per route-mile.

Under significant economies of density, the cost-minimizing market structure for a given route might call for a single firm, i.e., the route would be a natural monopoly. In the absence of any other scale economies, the national railway system could be made up of a large number of small firms, each with a local monopoly. Alternatively, if there were substantial economies of firm size without economies of traffic density, it would be economic to have a number of integrated nationwide railroads that competed on all their routes. However, with economies of density, and with economies of scope, and with some

---

[18] See Harris (1977).

economies of end-to-end long hauls, the cost effective structure of the rail industry is likely to be characterized by very few firms.

A.5 Empirical Evidence on Scale Economies

There are at least two approaches to measuring cost-scale relationships in the rail industry. The first way is to employ the expertise of those who have intimate knowledge of railroad operations in ascertaining whether the costly inputs required to supply rail services must be expanded in proportion to accommodate expansions in the quantities of services provided. This is known as the *engineering approach.* The second approach is statistical cost analysis--econometrically estimating the relationship between railroad costs and the levels of rail services provided. In the railroad industry, there is no conflict between the conclusions reached from these two different approaches. Both indicate quite clearly that railroad operations are characterized by increasing returns to scale, and that consequently the recovery of railroad costs requires that prices exceed marginal costs.

The first approach has been followed by a long succession of industry observers who provided a knowledgeable overview of the details of how economies of scale arise in rail operations. First, economies are created for the system as a whole by operations which are directly common to all traffic. Prominent among these are economies in network planning and management. If network management and control (e.g., billing, payroll, system-wide insurance, and other housekeeping functions) involves a fixed cost regardless of network size (above a certain threshold), these costs will be spread over a larger user base in a larger integrated rail system. Similar integration economies arise in communications and dispatching activities, and from increases in work force specialization within the repair facilities of larger systems. Finally, large railroads benefit from capital raising and other pecuniary economies (e.g., price concessions from suppliers). Indeed, this appears to be one of the most persistent advantages of firm size, with small incremental capital cost savings enjoyed out to very large scales. However, the capital-raising economies of scale are also associated with real resource savings. Negotiating a loan or new stock issue or obtaining necessary regulatory clearances entails transaction costs, some portions of which are nearly fixed. Clearly, the larger the issue is, the lower those costs are per unit of capital raised.

Second, the integration of the railroad system permits economies which directly benefit some traffic and indirectly benefit other system activities. Most ancillary plant (e.g., storage and marshalling yards, sidings, switches, and fueling and repair stations) can be utilized by more and more shippers without causing a corresponding increase in the amount of investment required. A coal shipper might need a storage and marshalling yard to hold its cars until a trainload volume is accumulated. If a mine produces only 20 carloads a day and holds them until 100 cars are available, a yard that could store and switch 100 cars would be required. However, on an independent operation basis, only 20 percent of the

yard would be utilized in the first day, 40 percent in the second, 60 percent on the third, etc. Yet, a railroad that connected with more mines might receive 20 cars a day from five mines and send a trainload every day. It would still only need the 100-car yard, but would have five times as many cars to share in the coverage of the investment and operating costs of the yard.

Similarly, a full siding is necessary if one train a day will meet one other train coming in the opposite direction--but the same size siding would be necessary if four trains were meeting four trains at the same place. Crossing protection must be built and maintained in a densely populated area whether the railroad sends one train a day or three trains a day over the track at the crossing. The same is true for switches, fuelling stations, and all other fixed plant investment. Once the plant is installed, a railroad can utilize it far more heavily with very little additional fixed investment cost. Also, a train of 40 cars, needs a crew of the same size as one of 60 cars. The ability to marshall cars of different shippers into a larger train also cuts other operating expenses. The engine power necessary for a longer and heavier train is not commensurate with the additional cars which have been added.

The statistical or econometric approach to analyzing railroad economies of scale has also had a long history. This history is rife with academic controversy and with steadily improving research methods. For example, some econometric studies found no evidence of rail economies of scale because they were founded on arbitrary allocations of costs between freight and passenger services. Other investigators failed to distinguish economies of scale from possible economies stemming from the geographic extent of a railroad's operations. Such studies incorrectly conclude that increasing returns to scale are absent from a finding that railroads covering more territory do not necessarily enjoy lower costs per ton-mile of freight.

Recent econometric studies in the United States have avoided these pitfalls and warrant several important conclusions. First, most of the rail system operates subject to increasing returns to scale and has elements of natural monopoly, whether considered in a single-product or a multiproduct setting. Second, as Fig. A.1 indicates, unit costs decline sharply with density, but at some point between 25 and 40 million annual gross ton-miles per route-mile, depending on the commodity mix, the cost curve flattens out and a large part of the traffic in the system flows over this flat part. This represents the level of minimum efficient density, and one can think of this as the capacity of a single track between two points, the fundamental indivisibility in the rail cost structure. Higher traffic density can be served at approximately constant cost by adding segments of parallel track and signalling devices. Third, for very short-haul, terminal-oriented railroads, the long-run cost curve seems to flatten out much sooner (at under 2 million net ton-miles per route-mile). Fourth, there are considerable economies of longer hauls.

Overall, these studies establish the presence of substantial economies of scale in the freight operations of railroads. They indicate that pricing at short- and long-run marginal costs would recover

less than 80 percent of total long-run costs. Also, high density traffic seems to exhaust the economies of scale experienced at lower densities, but significant diseconomies of scale do not occur as densities grow larger. Consequently, since all railroads have relatively low density traffic on many segments, and since most traffic flows on low density track while it is gathered and distributed, rail services exhibit substantial economics of scale overall. As a result, prices set at marginal costs would leave uncovered a substantial portion of total efficient railroad
costs.

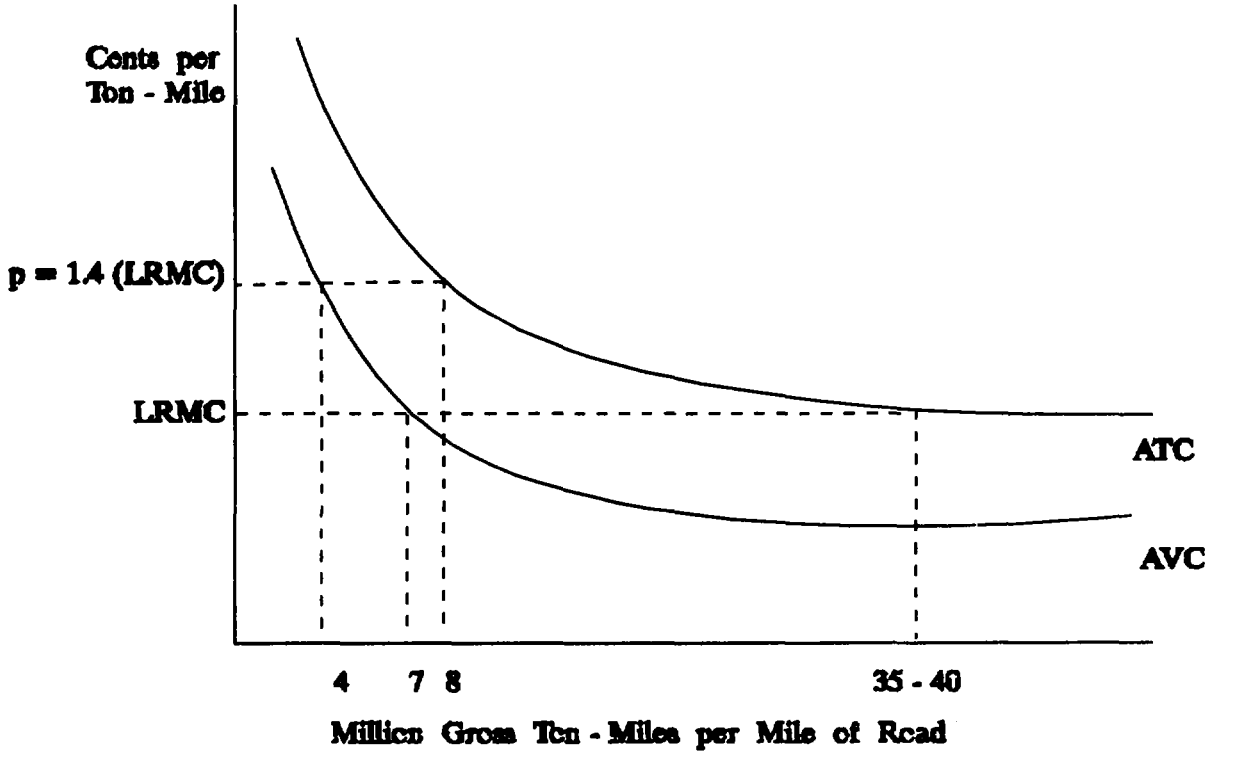

Figure A.1 Unit Total Cost, Operating Costs, and Traffic Density. Source: Levin (1981).

Appendix B. RAIL COSTS, PROFITABILITY, AND INFRASTRUCTURE CHANGES

Most of the statistical and econometric studies estimating rail costs and production functions suffer from two fundamental weaknesses. First, they generally fail to differentiate between way-and-structures capital, which is a measure of the quantity and quality of the capital utilized in the roadbed, and track, which in addition to being a proxy for the roadbed capital is also a measure of common carrier obligations to haul commodities. Second, they generally fail to take into account the effect upon costs of the route network and differentiate between high-density, fully utilized track and light-density, underutilized track.[19]

Way-and-structures (WS) capital is a measure of the capital utilized in the roadbed and as such should be treated as a conventional factor of production. An increase in the fixed factor, WS capital, should lead to a reduction in other factors and hence reduction in variable costs. In contrast, track and low-density track should be treated as technological variables that affect the costs of the railroad firm in a way that is not necessarily associated with conventional production theory. An increase in low-density route-miles or total track represents an increase in common carrier obligations and should therefore be associated with increases in expenditures on other factors of production.

A ceteris paribus reduction in WS capital will reduce the quality of the existing track and hence lead to cost increases by requiring increased amounts of variable factors--more money must be spent on equipment maintenance and train crews as the quality of the roadbed deteriorates and speeds are reduced. Similarly, a ceteris paribus reduction in track will not only be correlated with reduction in common carrier obligations and improvements in the quality of the existing track, but also with increases in its utilization. The first two considerations will tend to reduce costs, while the latter will tend to increase them, making the impact of reduced track somewhat ambiguous. Reduction in low-density track, on the other hand, will reduce common carrier obligations and their associated costs and will therefore tend to generate cost savings.

B.1 Railroad Costs and Infrastructure Variables

To assess the possible savings that would accrue from policies aimed at changing the railroad infrastructure, it is important to quantify the impact of changing the three main infrastructure variables--the amount of WS capital, general track, and low-density track--upon rail costs.

Ceteris paribus increases in WS capital will raise the amount of capital embodied in each mile of track and thus lead to reductions in variable costs. Indeed, econometric estimates (from the United States)

---

[19] See Friedlaender and Spady (1980).

reveal that a 10-percent increase in WS capital leads to over 4-percent decrease in variable costs consisting of approximately 11-percent savings in equipment usage, a 3-percent decrease in general labor, a 3-percent decrease in yard and switching labor, a 2-percent decrease in on-train labor, and a 0.6-percent savings in fuel and materials.[20] These estimates, therefore, seem to indicate that the main effect of an increase in WS capital is to decrease equipment requirements with somewhat lesser savings in the labor categories. This confirms the intuition that the savings in variable costs that result from an increase in WS capital have increased train speeds as their source.

Ceteris paribus reductions in light-density track are correlated with increases in the amount of capital embodied per mile of track and reductions in the proportion of low-density mileage; both of these factors should be associated with cost reductions. Econometric estimates indicate that a 10-percent reduction of low-density route-mileage would reduce total variable costs by approximately 3 percent. This comes about by reducing yard and switching labor costs by somewhat over 4 percent, reducing general labor and equipment expenditures by somewhat over 3 percent, and by reducing fuel and materials expenditures by less than 1 percent. Thus the primary savings arising from abandonment of low-density line are concentrated in transportation and switching categories associated with moving trains over lightly utilized track.

Finally, ceteris paribus reductions in general track are not only correlated with increases in capital embodied per mile of track, but are also correlated with increases in the proportion of low-density track. While the first factor should tend to reduce costs, the second should increase them. Econometric estimates reveal that a 10-percent reduction in general track or route-miles only leads to a reduction of total costs of less than 1 percent. In terms of factor utilization, reductions in general route-miles lead to sizeable reductions in equipment and materials expenditures, but increases in labor expenditures. Thus as the same volume of traffic is moved over a smaller network, increased expenditures on labor and switching are required, while savings on fuel and equipment are achieved.

B.2 Low Density Lines and Profitability

Rail costs are quite sensitive to changes in WS capital and light-density route-miles, but not to changes in general route-miles. A change in general track or route-miles without a concomitant change in low-density route-miles has a small impact on variable costs, but a significant effect on factor intensities. What distinguishes the provision of low-density service from that of general network expansion is the greater labor intensity of the former. Thus, efforts to adjust amounts of WS capital

[20] See Friedlaender and Spady, op. cit.

through roadbed maintenance or efforts to abandon light-density lines are likely to have a rather large impact on costs, while the abandonment of general track per se will lead to relatively few economies.

Econometric estimates from the United States reveal quite clearly that the low-density lines are a significant drain on railroad profitability, and constitute a serious impediment to the attainment of static and dynamic efficiency in the industry. The avoidable losses recoverable by abandonment appear to be quite significant. In addition, the burden of excess capacity seems to have a dynamic impact on efficiency. The abandonment of low-density lines stimulates the formation of new capital on the high-density portions of the rail network. First, since abandonment reduces the need for cross-subsidization, rates on the high-density lines are permitted to fall toward marginal cost. The lower rates attract additional traffic, and thus raise the level of desired capital. Second, abandonment of low-density lines lowers the cost of capital to rail firms by improving their long-run profitability and reducing the risk of bankruptcy.

REFERENCES

Bailey, Elizabeth E., "Contestability and the Design of Regulatory and Antitrust Policy," *American Economic Review,* 71 (1981): 178-183.

Baumol, William J., and Robert D. Willig, "Fixed Costs, Sunk Costs, Entry Barriers, and Sustainability of Monopoly," *Quarterly Journal of Economics*, 95 (1981): 405-31.

Baumol, William J., Panzar, John C. and Robert D. Willig, Contestable Markets and the Theory of Industry Structure, San Diego: Harcourt Brace Jovanovich, 1988.

Baumol, William J., and J. Gregory Sidak, "The Pricing of Inputs Sold to Competitors," *Yale Journal on Regulation*, 11 (1994): 171-202.

Boyer, Kenneth D., "The Costs of Price Regulation: Lessons from Rail Regulation," *Rand Journal of Economics*, 18 (1987): 408-16.

Braeutigan, Ronald R., "An Analysis of Fully Distributed Cost Pricing in Regulated Industries," *Bell Journal of Economics*, (1977): 182-195.

Braeutigan, Ronald R., "Optimal Pricing with Intermodal Competition," *American Economic Review*, 69 (1979): 38-49.

Braeutigan, Ronald R., "Socially Optimal Pricing with Rivalry and Economies of Scale," *The Rand Journal of Economics*, 15 (1984): 127-135.

Friedlaender, Ann F., "The Social Costs of Regulating the Railroads," *American Economic Review,* 61 (1971): 226-234.

Friedlaender, Ann F., The Dilemma of Freight Transport Regulation, Washington, D.C.: The Brookings Institution, 1969.

Friedlaender, Ann F., and Richard H. Spady, Freight Transport Regulation, Cambridge: The MIT Press, 1980.

Harris, R.G., "Economies of Traffic Density in the Rail Freight Industry," *Bell Journal of Economics,* 8 (1977): 556-564.

Kahn, Alfred E., The Economics of Regulation, Cambridge: The MIT Press, 1988.

Keeler, Theodore E., Railroad, Freight and Public Policy, Washington, D.C.: The Brookings Institution, 1983.

Levin, Richard C., "Allocation in Surface Freight Transportation: Does Rate Regulation Matter?" *Bell Journal of Economics*, 9 (1978): 18-45.

Levin, Richard C., "Railroad Rates, Profitability, and Welfare Under Deregulation," *Bell Journal of Economics*, 12 (1981): 1-26.

Levin, Richard C., "Exit Barriers and Railroad Investment," In Gary From, ed., Studies in Public Regulation, Cambridge: The MIT Press, 1981.

Moyer, Neil E., and Louis S. Thompson, "Options for Reshaping the Railway," The World Bank, WPS 926 (1992).

Ordover, Janusz, and Robert D. Willing, "An Economic Definition of Predation: Pricing and Product Innovation," *Yale Law Journal*, 90 (1981): 1-44.

Ordover, Janusz, Sykes, A., and Robert D. Willig, "Non-Price Anticompetitive Behavior by Dominant Firms Toward the Producers of Complementary Products," *Antitrust and Regulation*, The MIT Press, 1985.

Tye, William B., "The Voluntary Negotiations Approach to Rail Competitive Access in the Transition to Deregulation," *The Antitrust Bulletin*, (1987): 415-50.

Willig, Robert D., and William J. Baumol, "Using Competition as a Guide," *Regulation*, 1 (1987): 28-35.

Willig, Robert D., "Public versus Regulated Private Enterprise," The World Bank, Annual Bank Conference on Development Economics, 1993.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1491	Equilibrium Incentives for Adopting Cleaner Technology Under Emissions Pricing	Peter W. Kennedy Benoit Laplante	August 1995	E. Schaper 33457
WPS1492	Trade Policies, Macroeconomic Adjustment, and Manufactured Exports: The Latin American Experience	Sarath Rajapatirana	August 1995	J. Troncoso 37826
WPS1493	Migration and the Skill Composition of the Labor Force: The Impact of Trade Liberalization in Developing Countries	Ramón López Maurice Schiff	August 1995	J. Ngaine 37947
WPS1494	Adjustment and Poverty in Mexican Agriculture: How Farmers' Wealth Affects Supply Response	Ramón López John Nash Julie Stanton	August 1995	J. Ngaine 37947
WPS1495	Raising Household Energy Prices in Poland: Who Gains? Who Loses?	Caroline L. Freund Christine I. Wallich	August 1995	G. Langton 38392
WPS1496	Reviving Project Appraisal at the World Bank	Shantayanan Devarajan Lyn Squire Sethaput Suthiwart-Narueput	August 1995	C. Bernardo 37699
WPS1497	Public Choices between Lifesaving Programs: How Inportant are Lives Saved?	Maureen L. Cropper Uma Subramanian	August 1995	A. Maranon 39074
WPS1498	Decentralized Rural Development and Enhanced Community Participation: A Case Study from Northeast Brazil	Johan van Zyl Tulio Barbosa Andrew N. Parker Loretta Sonn	August 1995	M. Williams 87297
WPS1499	The Dynamics of Poverty: Why Some People Escape from Poverty and Others Don't—An African Case Study	Christiaan Grootaert Ravi Kanbur Gi-Taik Oh	August 1995	A. Sachdeva 82717
WPS1500	Agricultural Trade Liberalization in the Uruguay Round: One Step Forward, One Step Back?	Merlinda D. Ingco	August 1995	J. Ngaine 37947
WPS1501	Are Partner-Country Statistics Useful for Estimating "Missing" Trade Data?	Alexander J. Yeats	August 1995	J. Ngaine 37947
WPS1502	Active Labor Market Policies in the OECD and in Selected Transition Economies	Hartmut Lehmann	August 1995	WDR 31393

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1503	Africa's Growth Tragedy: A Retrospective, 1960–89	William Easterly Ross Levine	August 1995	R. Martin 39120
WPS1504	Savings and Education: A Life-Cycle Model Applied to a Panel of 74 Countries	Jacques Morisset César Revoredo	August 1995	N. Cuellar 37892
WPS1505	The Cross-Section of Stock Returns: Evidence from Emerging Markets	Stijn Claessens Susmita Dasgupta Jack Glen	September 1995	M. Davis 39620
WPS1506	Restructuring Regulation of the Rail Industry for the Public Interest	Ioannis N. Kessides Robert D. Willig	September 1995	J. Dytang 37161